## REMARKS

This is a Communication to submit two chapters from the protocols and review book titled DNA Microarrays: Gene Expression Applications -- Principles And Practice [Bertrand R. Jordan (Editor), Springer-Verlag, Berlin, Heidelberg, New York, 2001, 140 pages]. This information is believed to be significant in that the various authors in this book describe the very elements of DNA arrays that are the subject of the present invention.

In Chapter 1 of this book titled "DNA Arrays for Expression Measurement: An Historical Perspective," Dr. Jordan describes the history leading to the development of DNA arrays. On page 3, the author writes:

> **1**
> **Introduction**
>
> Basically, ***DNA arrays consist of a series of DNA segments regularly arranged on some kind of support***, and ***the expression measurement involves hybridising the whole array with a labelled DNA or RNA sample***. The essential feature is parallel processing: in a single experiment, information is obtained on each of the hundreds or thousands of entities present on the array (e.g. whether or not, and at which quantitative level, they hybridize with a given nucleic acid species). It is this parallelism that makes them so important at a time when many megabases of genome sequence and thousands upon thousands of genes need to be analysed. [emphasis added]

In the very next section, further developments on the "parallel processing" concept are covered:

> 2
>
> **The Forerunners: Colony Filters and Dot Blots**
>
> This principle of parallel processing was already implemented in the 1970s. ***Colony hybridisation (Grunstein and Hogness 1975) was used to search for specific genes among libraries; dot blots (Kafatos et al. 1979) and slot blots allowed homology determination or expression analysis on series of samples, with radioactive labelling in almost all cases (Fig. 1.1).*** The ***density*** of colony filters could be ***quite high, up to 10,000 on a Petri dish-sized membrane, but these colonies were arranged at random on the membrane*** since they resulted from plaque lifts or from direct spreading of a transformation reaction onto the membranes. Thus, ***no permanent record of the colonies existed***, and the experiments were directed at isolating one or several "positive" clones, ***the rest being discarded***. ***Dot and slot blots***, on the other hand, were done in ***ordered format***, often with 96-trough devices geared to microtitre plates, the ***DNA (or RNA) solutions were passed through the membrane under conditions conducive to binding***, and the resulting microtitre-sized dot blot displaying up to 96 spots was used for expression or homology analysis (Fig. 1.1). [emphasis added]

In the next chapter titled "Expression Profiling with cDNA Microarrays: A User's Perspective and Guide," (pages 13-33), Drs. Sean Grimmond and Andy Greenfield delve into further details on microarrays and their use. In a telling description on page 15 (Section 2), the authors write:

2

**Expression Profiling with cDNA Microarrays: The Basics**

The ***use of cDNA microarrays involves three stages*** that are summarized in Fig. 2.1. The ***first stage*** (Fig. 2.1A) involves the ***preparation, arraying and attachment of DNA probes (also known as elements) to a non-porous substrate***.[2] The DNA elements used to make expression microarrays are normally PCR products amplified from cDNA, using either gene- or vector-specific oligonucleotides. ***The non-porous substrate onto which the DNA probes are arrayed is typically a treated glass slide whose surface has been modified to bind DNA (traditionally coated with poly-L-lysine***[3] (Schena et al. 1995). A robotic arrayer is used to print out the DNA elements at very high density and the arrayed DNA is fixed to the surface.

The second stage of expression profiling involves preparation of labeled cDNA pools (known as labelled target) from a test and reference RNA sample. Each sample is labelled using a differently ***fluorescently labelled nucleotide***[4] (e.g. Cy5-dCTP) for reference. Cy3-

---

[2] **Presently pending claims 1576 and 1670 are directed to the following array subject matter:**

**1576. An *array* comprising a *non-porous substrate* having surfaces, each surface comprising *at least one double-stranded nucleic acid fixed or immobilized to one or more reactive groups or binding sites on said surface*, wherein at least one nucleic acid strand or a sequence therefrom comprises one or more non-radioactive chemical labels which comprise a non-radioactive signaling moiety or moieties which are quantifiable or detectable, wherein at least one nucleic acid strand or a sequence therefrom in one of said surfaces is different from at least one other nucleic acid strand or a sequence therefrom in another surface, and wherein said non-porous substrate comprises siliceous matter or polymeric material.**

**1670. An *array* comprising a *non-porous substrate* having surfaces, each surface comprising *at least one nucleic acid strand fixed or immobilized to one or more reactive groups or binding sites on said surface*, wherein at least one nucleic acid strand or a sequence therefrom in one of said surfaces is different from at least one other nucleic acid strand or a sequence therefrom in another surface, and wherein said non-porous substrate comprises siliceous matter or polymeric material.**

[3] **Several dependent pending array claims recite "polylysine (PPL)," including claims 1592-1593, 1597, 1686-1687 and 1691. Poly-L-lysine is a polyamino compound which is also covered by the terms "amine providing compound," "amine," "polyamin ," and related language. The latter terms are recited in claims 1590-1593, 1595-1598, 1684-1687 and 1689-1691.**

[4] **Several dependent pending array claims are directed to fluoresc ntly labeled nucleic acid. These claims includ claims 1634, 1636, 1726 and 1728 ("wherein said indicat r molecules are . . . a fluorescent c mpound . . ."); claims 1640, 1647, 1732 and 1739 ("wherein said directly produc d**

> dCTP for test RNA). Both labelled populations are then pooled and co-hybridised to the same cDNA arrays.
>
> After hybridisation and washing, the ***third stage*** involves ***quantifying[5] the test and reference signals of each fluorophore for each element on the array***, traditionally achieved by confocal laser scanning. Image analysis software is used to determine signal for each and the differentially expressed genes are identified.
>
> . . . [emphasis and footnotes added]

Later in the same chapter, on page 24, Drs. Grimmond and Greenfield elaborate on the chemistry exploited to attach DNA to non-porous substrates:

---

**non-radioactive signaling moiety or moieties comprise . . . a fluorogen . . ."); claims 1651 and 1743 ("wherein said enzyme or enzymatic reaction fluoresces . . ."); claims 1653, 1654, 1745 and 1746 ("wherein said non-radioactive signaling moiety or moieties are . . . a fluorescent compound . . ."); claims 1655, 1656, 1747 and 1748 ("wherein a non-radioactive signal is generated or generatable from said non-radioactive chemical label or labels by a means . . . fluorescent means . . ."); claims 1657, 1658, 1749 and 1750 ("wherein said non-radioactive chemical label or labels are . . . a fluorescent compound"); and claims 1663, 1664, 1755 and 1756 ("wherein said photom tric means are flu rometric techniques . . .").**

**[5] The term "quantifiable" is r cited in claim 1576 (" . . . wherein at least n nucl ic acid strand or a s quence therefrom comprises one or m r non-radioactive chemical labels which comprise a non-radioactive signaling moiety or moieties which ar *quantifiable* or det ctable, . . .").**

3.2.4.2
**Slide Chemistry**

Slide chemistry consists of ***the way DNA is attached to the non-porous glass substrate***[6] ***and the subsequent inactivation of the substrate-post-arraying. When this is done efficiently, DNA elements are successfully bound to the glass surface and then inactivated to prevent labelled cDNA targets from binding to the substrate during the hybridisation***[7]***, resulting in undesirable background.*** The sensitivity of this technology derives from the ***ability to detect weak fluorescent signals at a given element on a non-porous substrate in comparison to the very low background surrounding the elements.***

No matter what arraying system is being used, it is important to determine the best substrate and binding chemistry for a given objective. Some chemistries require the use of modified 5' amino-linked primers in the PCR amplification used to generate probe. Traditionally, ***poly-L-lysine-***[8] ***or silane-coated slides***[9] have been used. The most popular to date has been poly-L-lysine slides prepared "in house". The advantages of the poly-L-lysine chemistry are that it requires no DNA modification, it is extremely cheap and, once perfected, it provides a highly consistent performance. The problems associated with home manufacture originate from batch variation in the product, the shelf-life and the poorly understood maturing process that is required between slide coating and microarraying. Commercial poly-L-lysine-coated slides have tended to be substandard for microarray purposes because they were not designed or packaged for this purpose.

More recently, there has been a dramatic increase in the number of commercially available substrates for microarraying. These substrates are generally made using superior glass of uniform thickness and both monolayer (e.g. Corning, Telechem) and branched

---

**[6] See footnote 2 above.**

**[7] Claims 1599 and 1600 are directed to "blocking agents." The former claim recites "wherein said surface or surfaces have been treated with a blocking agent," and the latter recites that "said bl cking ag nt comprises D nhardt's s luti n."**

**[8] See footn te 3 above.**

**[9] Silane modifications to a n n-por us substrate is described in the specification (Example 1, pages 15-16). The compound, γ-aminopr pyltriethoxysilan , which provides a silane coating to the non-porous substrate, is recited in claims 1592, 1597, 1686 and 1691."**

> polymer (e.g. 3D Surmodics) substrates exist. Many of the newer substrates require the use of 5' amino-modified amplimers for element preparation. [emphasis and footnotes added]

Copies of chapters 1 and 2 from Jordan's DNA Microarrays book are attached to this paper as Exhibit 1.

The above-quoted portions of this book show that the basic elements of array technology -- as practiced today -- are found in Applicants' disclosed examples and claimed invention.

Favorable action on this application is respectfully requested.

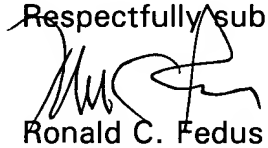
* * * * * * *

## SUMMARY AND CONCLUSIONS

No fee is believed due in connection with the filing of this Communication, the purpose of which is to present some new information that has recently come to the attention of Applicants' attorney. In the event that any other fee or fees are due, however, The Patent and Trademark Office is hereby authorized to charge the amount of any such fee(s) to Deposit Account No. 05-1135, or to credit any overpayment thereto.

If a telephone conversation would further the prosecution of the present application, Applicants' undersigned attorney request that he be contacted at the number provided below.

Respectfully submitted,



Ronald C. Fedus
Registration No. 32,567
Attorney for Applicants

**ENZO LIFE SCIENCES, INC.**
**(formerly named Enzo Diagnostics, Inc.)**
**c/o ENZO BIOCHEM, INC.**
**527 Madison Avenue, 9th Floor**
**New York, New York 10022**
**Telephone: (212) 583-0100**
**Facsimile: (212) 583-0150**

# PRINCIPLES AND PRACTICE

Springer
*Berlin*
*Heidelberg*
*New York*
*Barcelona*
*Hong Kong*
*London*
*Milan*
*Paris*
*Singapore*
*Tokyo*

B. R. Jordan (Ed.)

# DNA Microarrays: Gene Expression Applications

With 20 Figures

Springer

Dr. B.R. Jordan
Marseille-Génopole
Parc Scientifique de Luminy, Case 901
13288 Marseille, Cedex 9
France

ISBN 3-540-41507-6
Springer-Verlag Berlin Heidelberg New York

Library of Congress Cataloging-in-Publication Data
DNA microarrays: gene expression applications / Bertrand Jordan, ed. p. cm
Includes bibliographical references and index.
ISBN 3540415076 - ISBN 3540415084 (pbk.)
1. DNA microarrays-Laboratory manuals. 2. Gene expression-Laboratory manuals
I. Jordan, Bertrand. QP624.5.D726 D634 2001 572.8'65-dc21



Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de


Printed in Germany



Production: PRO EDIT GmbH, 69126 Heidelberg, Germany
Cover design: D&P, 69121 Heidelberg, Germany
Typesetting: AM-productions GmbH, 69168 Wiesloch, Germany
Printed on acid-free paper SPIN 10792609 39/3130Re - 5 4 3 2 1 0

# Preface

The aim of this book is to provide in compact form a comprehensive and practical survey of expression measurement using DNA arrays. I have endeavoured to assemble chapters written by scientists who are actual users of the technology and have had to cope with the various practical problems involved in setting up new methods in the laboratory; I believe this is how such a book can be most useful to its readers.

Chapter 1 provides some background on the history of DNA array development; Chapters 2, 3 and 4 focus on various types of microarrays: glass microarrays are described by experienced academic users, while the less-known (but in some situations superior) nylon microarrays are presented by their developers together with some information on nylon macroarrays that are still widely used. Chapter 5 describes in detail the use of oligonucleotide chips in a research laboratory, again with emphasis on practical aspects. The principles and practice of both data acquisition and data mining in the field of expression measurement are presented in Chapter 6 by very experienced authors, together with a wealth of Internet resources that are particularly useful in this fast-moving field. A short last chapter (Chapter 7) attempts to forecast the likely evolution of this field.

All the authors have strived for clarity and insistence on practical aspects; I can only hope that the result will be satisfactory for the reader.

Marseille-Génopole, March 2001 B. R. Jordan

P.S. One vexing point in the terminology of DNA arrays is that the DNA segments bound to the device are called "target" by some and "probe" by others, while the reverse applies to the labelled material prepared from the sample. I have not tried to force a solution to this problem, apart from ensuring consistency within each chapter.

# Contents

## Chapter 1



## Chapter 2

## Chapter 3



## Chapter 4





## Chapter 5



## Chapter 6



## Chapter 7

CHAPTER 1

# DNA Arrays for Expression Measurement: An Historical Perspective

BERTRAND R. JORDAN[1]



[1] Marseille-Génopole, Parc Scientifique de Luminy, Case 901, 13288 Marseille, Cedex 9, France

## 1 Introduction

Basically, DNA arrays consist of a series of DNA segments regularly arranged on some kind of support, and the expression measurement involves hybridising the whole array with a labelled DNA or RNA sample. The essential feature is parallel processing: in a single experiment, information is obtained on each of the hundreds or thousands of entities present on the array (e.g. whether or not, and at which quantitative level, they hybridise with a given nucleic acid species). It is this parallelism that makes them so important at a time when many megabases of genome sequence and thousands upon thousands of genes need to be analysed.

## 2 The Forerunners: Colony Filters and Dot Blots

This principle of parallel processing was already implemented in the 1970s. Colony hybridisation (Grunstein and Hogness 1975) was used to search for specific genes among libraries; dot blots (Kafatos et al. 1979) and slot blots allowed homology determination or expression analysis on series of samples, with radioactive labelling in almost all cases (Fig. 1.1). The density of colony filters could be quite high, up to 10,000 on a Petri dish-sized membrane, but these colonies were arranged at random on the membrane since they resulted from plaque lifts or from direct spreading of a transformation reaction onto the membranes. Thus no permanent record of the colonies existed, and the experiments were directed at isolating one or several "positive" clones, the rest being discarded. Dot and slot blots, on the other hand, were done in ordered format, often with 96-trough devices geared to microtitre plates; the DNA (or RNA) solutions were passed through the membrane under conditions conductive to binding, and the resulting microtitre-sized dot blot displaying up to 96 spots was used for expression or homology analysis (Fig. 1.1).

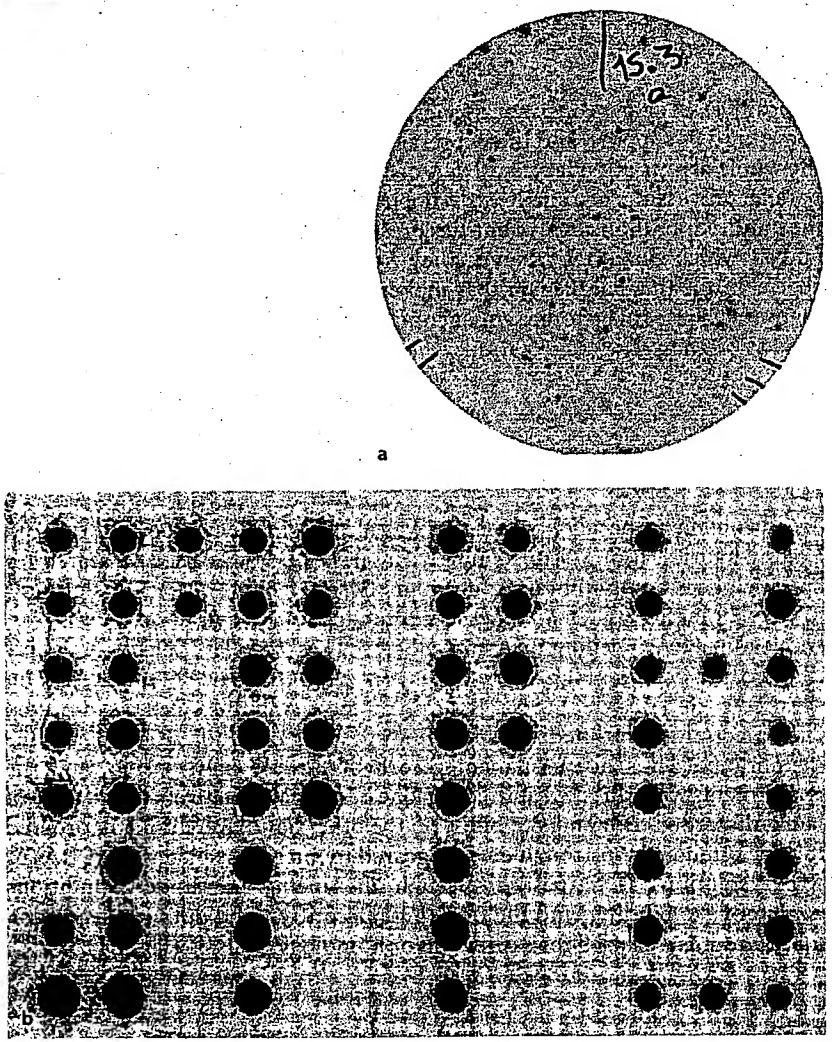



**Fig. 1.1. a** A phage plaque lift on round nitrocellulose membrane (approx. 1,000 plaques) with a few positive plaques showing up after hybridisation and exposure to X-ray film. Marks on edge of filter are used for subsequent critical positioning of the autoradiograph to recover "positive" plaque material from the Petri dish, since plaques are randomly distributed over the surface. Diameter is approximately 8 cm. **b** An example of the first implementations of ordered DNA arrays, a dot-blot performed with a 96-well filtration device and hybridised with a radioactive probe. Example shown is a screening application with positive/negative samples. Dimensions are ca. 8×12 cm



## 3
## "High-Density" Filters

A major change in the field was the development in the late 1980s of robotic devices ("gridding robots") that made it possible to spot bacterial colonies in a compact and regular pattern. The resulting "high-density filters" contained approximately 10,000 spots on a 22×22 $cm^2$ surface, corresponding to a "pitch" (centre to centre spacing) of approximately 2 mm (Fig. 1.2). Of course, these colonies had to be ordered previously , i.e. picked from agar plates and distributed into microtitre plates for storage and filter construction. This was often done by hand, although "picking robots" were being developed and reliable commercial models appeared on the market in the early 1990s.

Hans Lehrach's group, at EMBL (Heidelberg, Germany), then at ICRF (London, UK) and now at the Berlin Max Planck Institute, was the major proponent of this approach for genome analysis (Hoheisel et al. 1991; Lennon and Lehrach 1991), at a time when the approach taken in the USA relied almost exclusively on PCR methods and PCR screening of cleverly arranged pools of clones (Olson et al. 1989; Green and Olson 1990) - hybridisation being considered "not robust enough" for systematic genome work. Lehrach and coworkers introduced the concept of the "reference library" as a defined and more or less permanent set of clones, each of which has a definite physical address (e.g. well B12 of plate 312), is present at a known position in a set of high-density membranes, and is stored with all the information obtained in a relational database (Nizetic et al. 1991). This made it possible to acquire, store and correlate data from many different laboratories, as long as they used the same membranes and sent back their results to the central database, a major hurdle in practice.

For a while, high-density membranes were mostly used for library access. Several resource centres accumulated various genomic and cDNA libraries and distributed them to laboratories in the form of high-density membranes. When the user had identified a positive spot using a probe for a gene of interest, the centre then provided the corresponding clone (Fig. 1.2). This approach worked fairly well, at a time when sequence information was very scarce, and allowed many users to progress much faster with their projects. As noted before, information feedback to the resource centre has always proved difficult to enforce; the data coming from very diverse laboratories also turned out to be difficult to handle, so that, contrary to early expectations, this approach did not contribute much to the construction of large-scale physical maps.

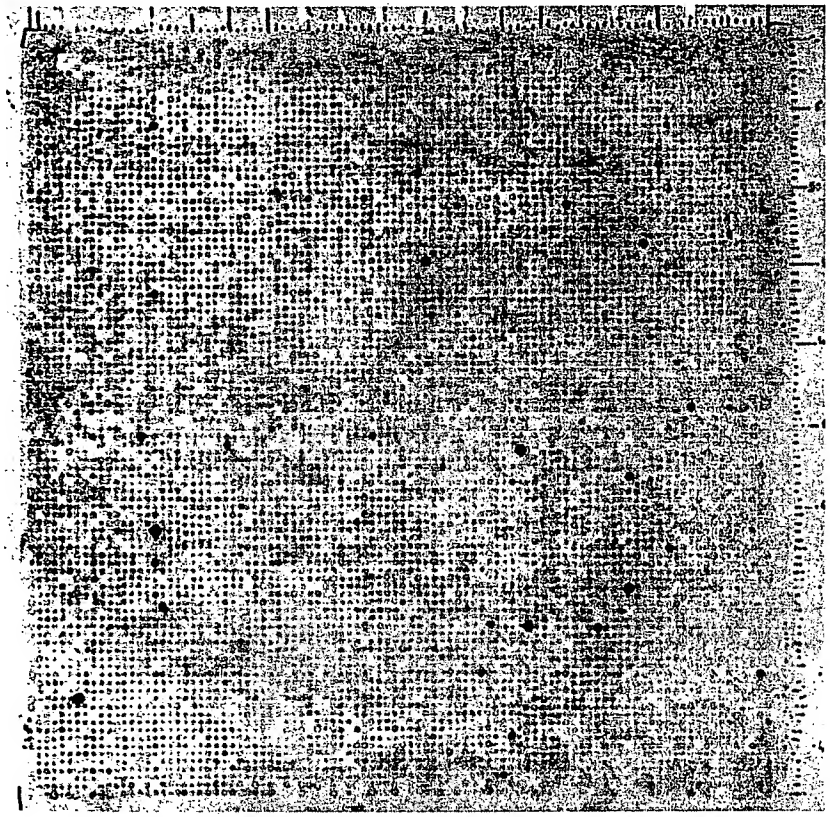


**Fig. 1.2.** An early "high-density filter" used to access genomic or cDNA libraries. In this case, 10,000 bacterial colonies containing DNA segments cloned in cosmids have been spotted and grown on a 22×22 cm nylon membrane. Hybridisation with a probe prepared from the insert of a cDNA clone reveals one or several positive spots, indicating the corresponding genomic clone; the relevant cosmid can be ordered from the resource centre that provided the membrane by indicating its coordinates. The whole grid is visible because of light background hybridisation; successive hybridisations to search for various genomic clones are performed on the same membrane without stripping (to avoid loss of material), hence the large number of "positive" spots. This was the first widely used application of high-density DNA arrays, pioneered by Hans Lehrach's group and involving only positive/negative scoring, usually by visual inspection



## 4
## Colony Filters and Expression Analysis (Qualitative)

The use of unordered or ordered colony filters containing cDNA clones for expression analysis began in the early 1980s with a qualitative application, differential screening. Parallel hybridisation of "identical" membranes containing clones from conventional or subtracted cDNA libraries with complex labelled cDNA mixtures prepared from two different samples was used, alone or in conjunction with other methods, to pinpoint genes whose expression was different under the two conditions. Using radioactivity and X-ray film detection, this method was necessarily qualitative; it was also cumbersome and suffered from a number of technical problems, but was nevertheless instrumental in isolating several important genes such as the T cell receptor (Hedrick et al. 1984) or the CTLA (cytotoxic T lymphocyte-associated transcripts) series of molecules (Brunet at al. 1988).

## 5
## Quantifying Expression with High-Density Filters

The concept of using the newly developed imaging plate systems that were just beginning to penetrate biological research laboratories (Amemiya and Miyahara 1988) for quantitative acquisition of these data and more refined analysis of expression patterns was fairly obvious and discussed as early as 1990, with a first publication reporting actual data in 1992 (Gress et al. 1992), but the full implementation of the technology took a fair amount of time. This is not unusual when moving from a qualitative to a quantitative application, especially as expression measurement with DNA arrays involves the quantification of very weak signals if data for genes expressed at very low levels are required. A number of artefacts leading to spurious data (especially with the unsequenced cDNA libraries of that period) had to be identified and taken care of (Nguyen et al. 1995), standardisation methods had to be worked out (Bernard et al. 1996), and the rather primitive and user-unfriendly image analysis programs of the period had to be improved (Granjeaud et al. 1996) – a task made easier by the quickly increasing computing power available to scientists. Automation of PCR and standardisation of plasmid vectors used for cDNA libraries led to a shift from colony filters to membranes on which amplified DNA has been deposited, although colony filters are still used in some cases.

During the first half of the 1990s, a few groups worked out the methods, published proof-of-principle papers (Nguyen et al. 1995; Zhao et al. 1995; Pietu et al. 1996) and began accumulating expression data in different systems (Fig. 1.3), leading later to biological results published in the relevant journals. Some of this work was performed in "gene discovery" mode, i.e. measuring the

expression level in a number of conditions for a large set of genes (represented by sequenced or, in some cases, unsequenced cDNAs), in order to pinpoint genes whose expression patterns suggest their involvement in particular biological events (see, e.g., Carrier et al. 1999). In this type of approach, extensive biological knowledge is necessary at the start, to define precisely the conditions that are likely to yield the most significant information, and later to choose the "best" genes out of the dozens that display the required expression profile and to define and perform the additional experiments that will allow this choice to be done most effectively. In this context, the biological samples are used to obtain information on the genes and to highlight the most interesting ones (preferably "new", i.e. not yet described in detail), whose close study is likely to be the most rewarding with regard to the biological question approached.

Expression measurement was also performed in an "expression profiling" mode, in which the set of genes (often more restricted) was chosen a priori and usually well known, and the objective was to obtain information on the samples: analysis of the expression profile for a series of tumours, for instance, in the hope of obtaining prognostic and therapeutic information (see, e.g., Bertucci et al. 1999b). In this case the genes are used as tools to derive information on the samples, exactly the opposite of the previous situation. Of course, if technical progress, miniaturisation and decreasing costs make it possible to routinely include all human genes in every experiment, the two approaches will eventually coalesce.

The high-density filter (under the more trendy name of "macroarray") is still with us: for experiments of moderate scope, it performs quite adequately and blends well with current laboratory practice and equipment. In fact, a number of manufacturers market such membranes with sets of cDNA clones (in the form of PCR products or bacterial colonies) from various organisms, an indication of the continuing popularity of the method. Even a firm like Incyte, which has heavily promoted the microarray approach after acquiring a specialist company, Synteni, now also sells macroarrays.



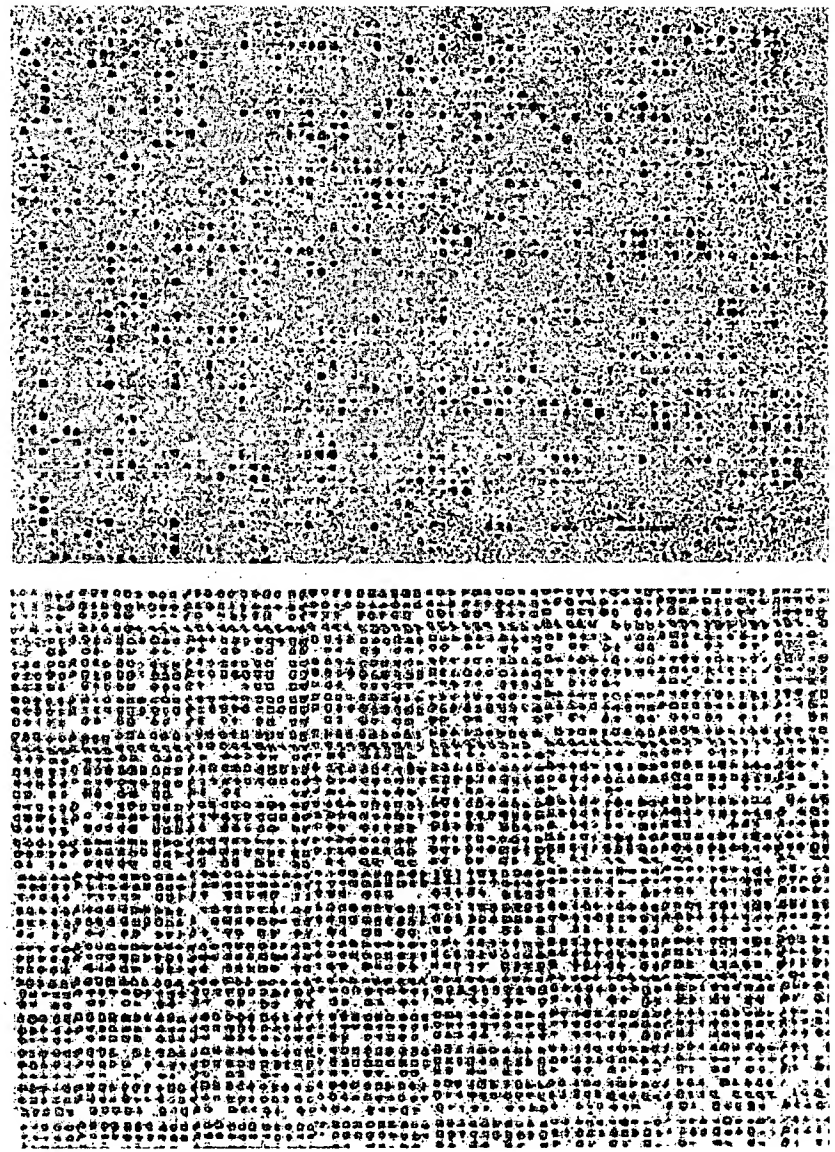


**Fig. 1.3.** Typical macroarray on nylon membrane, in this case measuring 8×12 cm and containing 3,072 bacterial colonies. *Bottom* hybridisation with a common vector sequence showing all spots (and allowing measurement of relative amount of DNA in each of them); *top* hybridisation with radioactive cDNA made from total mRNA. Hybridisation intensities are acquired and quantified using an imaging plate system, allowing measurement of relative mRNA abundances

## 6
## Miniaturisation: cDNA Microarrays

In the mid-1990s, miniaturisation became a major issue in the further development of DNA arrays, with the aim of increasing the number of genes assayed in a single experiment, and also of reducing sample usage – although most current systems still require microgram amounts of messenger RNA, a major limitation in practice. One avenue involved scaling down cDNA arrays to a pitch (centre to centre spot distance) of 200 to 400 μ (up to 2,500 spots/cm²). This was done using optical detection methods (fluorescence) because of their superior resolution, and depositing the DNA spots on very planar supports (glass slides) to allow intensity measurement with confocal optics in order to achieve the required sensitivity. First published in 1995 (Schena et al. 1995), this approach has blossomed and given rise to a number of important studies. The use of fluorescent probes allows dual labelling, simplifying comparisons and facilitating standardisation of series of experiments; very good sensitivity (in terms of detection of messenger RNAs expressed at low levels) has been obtained, although sample requirements remain high. Microarrays can be constructed in the laboratory; the necessary equipment is now commercially available, although the expense and logistics are not trivial. Ready-made arrays have now appeared on the market; this has been a relatively slow process due not only to the time taken to build up the necessary logistics but also to intellectual property issues. Generic microarrays should also be made available by academic resource centres. Chapter 2 gives an overview of this important approach and a concrete description of its implementation from a user's perspective.

Nylon microarrays can also be produced. Because of the intrinsic fluorescence of all nylon supports (so far), detection must be performed by enzymatic means that are convenient and affordable, but relatively insensitive (Chen et al. 1998), or with (33P) radioactive labelling, using high-resolution detectors that provide sufficient resolution to quantify arrays with 400-μ pitch. In this form the method makes expression profiling at reasonable sensitivity with very small biological samples (Bertucci et al. 1999a) possible. These two facets are described in Chapters 3 and 4.

## 7
## Miniaturisation: Oligonucleotide Chips

The other, competing approach is that of oligonucleotide chips, pioneered by S. Fodor and the firm Affymetrix. These glass chips, carrying hundreds of thousands of small (24×24 μ) "features", each containing several million copies of a given oligonucleotide (20- to 25-mer), were originally developed for "quasi-sequencing" (mutation detection) applications (Fodor et al. 1991).



They have been shown to allow expression measurement as well, at the expense of assaying each gene with several (20–40) oligonucleotides and controls, in order to average out signal and background artefacts due to the vagaries of short oligonucleotide hybridisation (Wodicka et al. 1997). Their manufacturing process, very similar to that of electronic devices, promises further miniaturisation beyond the present 400,000 feature chip. Being based solely on sequence knowledge, they do not require the cumbersome logistics of cDNA clone storage and PCR amplification, in contrast to cDNA arrays; however, the approach lacks flexibility, the chips are very expensive and their use in the academic sector is still rather limited. Chapter 5 provides a description of the practice of using these chips in an (industrial) research laboratory. Alternative approaches to oligonucleotide chips (notably different synthesis methods) may change the outlook and make this approach more flexible and user-friendly.

## 8
## Data Acquisition, Storage and Mining

The importance of software issues in expression measurement was not immediately recognised, but quickly became apparent as more and more massive data began to flow from these experiments. It includes a number of aspects, from verification of the validity of measurements to sophisticated data mining through data representation and storage issues. Although available computing power has increased dramatically in the last decade, these issues are still far from being satisfactorily solved. They are tackled in Chapter 6.

## 9
## A (Provisional) Conclusion

It is certain that expression measurement by DNA array methods will continue to be an important component of most genomics projects, for a very simple reason: it is, and will remain for some time, the only method that can add some functional information to thousands of sequences, with a throughput of the same order as that of DNA sequencing. However, it is not easy to forecast the directions in which this field will develop. The technology is still extremely fluid and new developments may have strong and short-term impact. In addition, as this field has been heavily invested by industry, commercial imperatives lead to scanty and overoptimistic communication which sometimes makes it difficult to foresee rationally future trends. I have, however, tried to sketch likely developments in the last chapter of this book (Chapter 7).

## References


Amemiya Y, Miyahara J (1988) Imaging plate illuminates many fields. Nature 336:89–90

Bernard, K, Auphan, N, Granjeaud, S, Victorero, G, Schmitt-Verhulst AM, Jordan BR, Nguyen C (1996) Multiplex messenger assay: simultaneous, quantitative measurement of expression for many genes in the context of T cell activation. Nucleic Acids Res 24:1435–1443

Bertucci F, Bernard K, Loriod B, Chang YC, Granjeaud S, Birnbaum D, Nguyen C, Peck K, Jordan BR (1999a) Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. Hum Mol Genet 8:1715–1722

Bertucci F, Van Hulst S, Bernard K, Loriod B, Granjeaud S, Tagett S, Starkey M, Nguyen C, Jordan B, Birnbaum D (1999b) Expression scanning of an array of growth control genes in human tumor cell lines. Oncogene 18:3905–3912

Brunet JF, Denizot F, Golstein P (1988) A differential molecular biology search for genes preferentially expressed in functional T lymphocytes: the CTLA genes. Immunol Rev 103:21–36

Carrier A, Nguyen C, Victorero G, Granjeaud S, Rocha D, Bernard K, Miazek A, Ferrier P, Malissen M, Naquet P, Malissen B, Jordan BR (1999) Differential gene expression in CD3e- and RAG-1 deficient thymuses: definition of a set of genes potentially involved in thymocyte maturation. Immunogenetics 50:255–270

Chen JJ, Wu R, Yang PC, Huang JY, Sher YP, Han MH, Kao WC, Lee PJ, Chiu TF, Chang F, Chu YW, Wu CW, Peck K (1998) Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. Genomics 51:313–324

Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D (1991) Light-directed, spatially addressable parallel chemical synthesis. Science 251:767–773

Granjeaud S, Nguyen C, Rocha D, Luton R, Jordan BR (1996) From hybridization image to numerical values: a practical, high throughput quantification system for high density filter hybridizations. Genet Anal Biomol Eng 12:151–162

Green ED, Olson MV (1990) Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. Proc Natl Acad Sci USA 87:1213–1217

Gress TM, Hoheisel JD, Lennon GG, Zehetner G, Lehrach H (1992) Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. Mamm Genome 3:609–619

Grunstein M, Hogness DS (1975) Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc Natl Acad Sci USA 72:3961–3965

Hedrick SM, Cohen DI, Nielsen EA, Davis MM (1984) Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature 14(308):149–153

Hoheisel JD, Lennon GG, Zehetner G, Lehrach H (1991) Use of high coverage reference libraries of *Drosophila melanogaster* for relational data analysis. A step towards mapping and sequencing of the genome. J Mol Biol 220:903–914

Kafatos FC, Jones CW, Efstratiadis A (1979) Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. Nucleic Acids Res 7:1541–1552

Lennon GG, Lehrach H (1991) Hybridization analyses of arrayed cDNA libraries. Trends Genet 7:314–317

Nguyen C, Rocha D, Granjeaud S, Baldit M, Bernard K, Naquet P, Jordan BR (1995) Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. Genomics 29:207–215

Nizetic D, Zehetner G, Monaco AP, Gellen L, Young BD, Lehrach H (1991) Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: their potential use as reference libraries. Proc Natl Acad Sci USA 88:3233–3237

Olson M, Hood L, Cantor C, Botstein D (1989) A common language for physical mapping of the human genome. Science 245:1434–1435

Pietu G, Alibert O, Guichard V, Lamy B, Bois F, Leroy E, Mariage-Sampson R, Houlgatte R, Soularue P, Auffray C (1996) Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. Genome Res 6:492–503

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470

Wodicka L, Dong H, Mittmann M, Ho MH, Lockhart DJ (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. Nat Biotechnol 15:1359–1367

Zhao N, Hashida H, Takahashi N, Misumi Y, Sakaki Y (1995) High-density cDNA filter analysis: a novel approach for large-scale, quantitative analysis of gene expression. Gene 156:207–213


CHAPTER 2

# Expression Profiling with cDNA Microarrays: A User's Perspective and Guide

SEAN GRIMMOND[1] and ANDY GREENFIELD[2]




[1] Institute of Molecular Bioscience, University of Queensland, St. Lucia, Queensland 4072, Australia

[2] MRC Mammalian Genetics Unit, Harwell, Did ot, Oxfordshire OX11 0RD, UK

## 1 Introduction

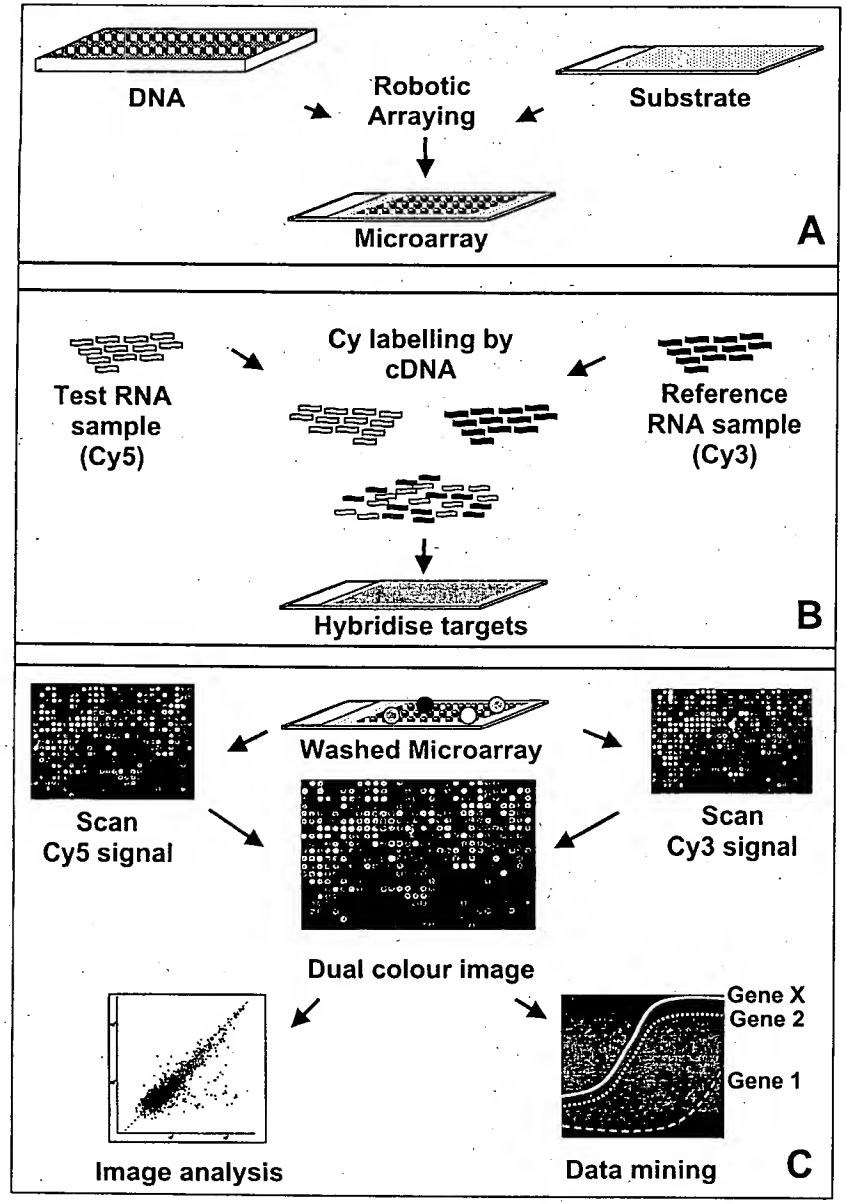
Genomic science has now advanced to the point where it is possible to define the genomic structure and gene content of any organism. Such advances have led to the development of expression tools that can study gene expression in a massively parallel fashion. These methods are also affordable, sensitive, discriminating, and require minimal sample RNA. cDNA microarray expression profiling is a rapidly developing technology that makes possible monitoring of gene expression on a genome-wide scale.

When RNA source is not limiting expression profiling can reach sensitivities of detection as high as 1/500,000 transcripts (Schena et al. 1996). The discriminating power of microarrays is similar to that of other hybridisation techniques (such as Southern or Northern hybridisations). When performed under stringent conditions it is possible to discriminate between elements that possess less than 85% nucleotide homology (DeRisi et al. 1996). Techniques have been established to make RNA target from minute amounts of RNA (Dixon et al. 1998; Lou et al. 2000) and arrays containing up to 18,000 elements have been reported (Alizadeh et al. 2000). The true power of this technology was demonstrated by expression analysis of the entire gene complement of *Saccharomyces cerevisiae* (DeRisi et al. 1997). The challenge now is to exploit DNA microarrays for expression profiling of more complex systems such as mammalian cell lines and tissue samples.

## 2 Expression Profiling with cDNA Microarrays: The Basics

The use of cDNA microarrays involves three stages that are summarised in Fig. 2.1. The first stage (Fig, 2.1A) involves the preparation, arraying and attachment of DNA probes (also known as elements) to a non-porous substrate. The DNA elements used to make expression microarrays are normally PCR products amplified from cDNA, using either gene- or vector-specific oligonucleotides. The non-porous substrate onto which the DNA probes are arrayed is typically a treated glass slide whose surface has been modified to bind DNA (traditionally coated with poly-L-lysine (Schena et al. 1995). A robotic arrayer is used to print out the DNA elements at very high density and the arrayed DNA is fixed to the surface.

The second stage of expression profiling involves preparation of labelled cDNA pools (known as labelled target) from a test and reference RNA sample. Each sample is labelled using a different fluorescently labelled nucleotide (e.g. Cy5-dCTP for reference, Cy3-dCTP for test RNA). Both labelled populations are then pooled and co-hybridised to the same cDNA array.






After hybridisation and washing, the third stage involves quantifying the test and reference signals of each fluorophore for each element on the array, traditionally achieved by confocal laser scanning. Image analysis software is used to determine signal for each and the differentially expressed genes are identified.

Even at this early stage in the development of microarray technology, many laboratories worldwide are looking to establish expression profiling facilities. The aim of the rest of this chapter is to describe the practical aspects of microarray experimental design and certain considerations worth making prior to establishing a DNA microarray expression system. The material in this chapter is based predominantly on our own experiences of constructing and using cDNA microarrays with commercially available hardware and software. A conventional description of our methodologies and data can be found in Grimmond et al. (2000).

## 3
## Production of cDNA Microarrays

### 3.1
### Experimental Design

Expression profiling using cDNA microarrays has the potential to identify small differences in gene expression levels (in the order of twofold) between any two biological samples. This level of sensitivity and resolution can create problems of interpretation if experiments are not designed carefully. The following are general considerations for any expression profiling experiment.

#### 3.1.1
#### Minimise Transcriptional Consequences Unrelated to the Biology

It is essential that the test and reference samples be carefully chosen to minimise the detection of differences that are unrelated to the biology of interest. It is important to exclude as much inter-sample variation as possible, from the culturing and sampling of the tissues to the preparation of RNA. This requires careful monitoring of reagents and plastic-ware to ensure that the same batches

**Fig. 2.1.** Three part schematic showing stages of an expression profiling experiment. A Microarray manufacture involves spotting of DNA elements onto specially coated glass slides using robotic delivery system. B Target labelling involves preparing fluorescently labelled cDNA pools from test and reference RNA samples and then hybridising both targets to same microarray. C After hybridisation, array is washed and then scanned with a confocal laser scan or CCD camera to determine amount of test and reference fluorescence hybridised to every element on the array. These data are often displayed as a dual colour image to allow for rapid identification of differentially expressed genes. Computer software is then used to correct true signal values and provide relative quantification levels of gene expression and can ultimately be mined using bio-informatic tools

are used throughout when inter-experiment comparison is required. In the case of the analysis of mouse models, genetic background differences between wild type and mutant mouse strains should be considered. It is unclear how many significant environmental factors may exist to confound a comparative analysis – the ambient temperature, humidity, time of day and individual researcher involved at the time of sampling may all affect the expression profiles generated.

### 3.1.2 Collect Appropriate Samples for Analyses

Firstly, it is always important to focus on target tissues that are central to the biological questions being addressed. Poor choice of tissue sample can result in little data of value, because critical transcriptional differences either are not present in the samples chosen or are diluted out by contaminating cell types. In the case of expression profiling of tumours, stromal cell contamination can vary widely between individual tumour samples. If total RNA is extracted from whole solid tumour samples and expression profiles compared, this variation in degree of stromal contamination will be reflected as expression differences. To avoid these sorts of problems, some researchers have resorted to collecting individual critical cell types from tissue samples via cell-sorting techniques such as laser capture microscopy (LCM) (Kacharmina et al. 1999).

When attempting to identify the transcriptional differences between reference and test samples, it is crucial that the test sample be collected at a suitable time-point, e.g. at a certain time after a particular challenge. (Challenges may range from a mutation to the administration of a particular pharmaceutical agent.) If samples are collected too soon, important biological responses may be missed and few transcriptional differences will be observed. If the time-point selected is too late, it is likely that many of the differences observed will be unrelated to the specific challenge but will simply reflect general consequences way downstream. To circumvent these problems, it is always preferable to perform a series of experiments at different time-points to assess expression profiles. Such a time-course of expression profiles also facilitates clustering of genes into particular functional groups that may shed light on the specific pathways affected by a challenge.

### 3.1.3 Use Elements Relevant to the Biology

Even if one manages to collect the most biologically relevant material to examine, the informativeness of the data generated is only as good as the elements that are interrogated. If the identification of genes activated in response to a given stimulus is the objective, elements generated from normal



resting tissues are unlikely to contain relevant probes. To this end many researchers have attempted to generate cDNA libraries which are most relevant to the systems they wish to study. Conventional cDNA libraries are not normally used without some adjustment of the level of abundant classes of elements by normalisation or subtraction (Soares et al. 1994; Bonaldo et al. 1996).

The recent availability of large minimal sets of cDNA clones (e.g. the RIKEN and DFKZ minimal cDNA sets and the commercially available sets (http://www.riken.go.jp/eng/index.html, http://www.dkfz-heidelberg.de/abt0840) has meant that a genomics-oriented approach is now possible. Although such sets are not biased towards a particular biological system, they have the advantage of providing minimal element redundancy. The logical conclusion of this approach is the complete genome set, and it is likely that these will be available in the not too distant future. The challenge will then be to increase probe density, allowing a complete genome set to be printed as a single array.

### 3.1.4 Always Reduce Complexity

Expression studies on cell lines are capable of detecting less than one transcript per cell when labelled target is not limiting. As the transcriptional complexity of an RNA sample increases, however, the overall sensitivity of an array experiment decreases. This becomes an issue when attempting to perform expression profiling on whole organs or tissue samples, but can be circumvented with the use of specific cell-capture methodologies allied to RNA amplification protocols (Kacharmina et al. 1999).

## 3.2 Microarray Construction

### 3.2.1 Practical Tips

A microarray experiment is only as good as the arrays used to perform the expression profiling. Extreme care should be taken when preparing both substrates and elements and the following tips highlight common sources of background on arrays. Wearing non-latex gloves and handling slides with forceps minimises exposure of substrates to oil and talc (common sources of non-specific background hybridisation). Solutions should be filtered to remove dust or particulate material. Frosted slides should be avoided as the white paint is often distributed across the surface of the substrate. Slides should always be spun dry at ambient temperature (500 rpm for 5 min) to prevent watermarks on the arrays. Watermarks can give local background that may not appear until hybridisation.

### 3.2.2 Elements

Elements can be divided into two classes: experimental and controls. In the case of cDNA microarrays, the experimental class of elements consists of PCR products generated from cDNAs whose expression profile is unknown. The number of these clones varies from study to study, but they usually comprise the bulk of elements on any given array.

As in all scientific experimentation, the control class of elements is essential to gauge the success and reliability of any given data set. In the case of establishing expression profiling, controls are vital to rapidly identifying problem areas in the methodology. Control elements can be broken down into three groups: (1) negative controls, (2) positive controls and (3) key controls.

1. The negative control class comprises a set of probes designed to detect non-specific and artefactual hybridisation signals. Typically we include a series of non-mammalian probes (plant or bacterial cDNAs) that should not detect any specific signal. A poly A tract probe is included to detect poor blocking of oligo dT tails found in labelled cDNA target, and a CoT1 DNA probe is included to detect non-specific signal associated with poor blocking of repeat sequences. Care should be taken to precipitate CoT1 DNA prior to arraying because commercial sources of CoT1 DNA supply the material in a Tris-buffered solution that prevents efficient attachment of the DNA to the treated glass substrate. Blank elements of arraying solution should also be included to confirm that probe carry-over is not taking place.
2. The positive controls used in our system were designed to evaluate the success of different stages of the expression profiling process. Firstly, fluorescently labelled cDNA probe was arrayed onto key positions of the array to act both as a positional guide and as a control to measure element retention throughout the entire procedure, from array spotting to post-hybridisation washes. Figure 2.2 shows the tracking of element retention on poly-L-lysine-coated slides by the use of printed fluorescently labelled probes. One disadvantage of this sort of control element is that it is prone to bleaching over time. It is important to note that the arraying of fluorescently labelled PCR products is not recommended for arraying systems that use quill-type pins because probe material can coat the pin and result in carry-over. All our arraying was performed using a GMS/Affymetrix 417 microarrayer that uses a solid pin and ring system for spotting. Elements derived from highly expressed mammalian genes (e.g. *Hprt*) can also be used as positional reference markers in key locations on the array (e.g. in each corner of the array) as they ensure a strong signal after hybridisation.
- Labelling efficiency was monitored by the addition of known quantities (50 and 5 ng) of in vitro transcribed (IVT) RNA from two plant cDNAs to each RNA sample prior to labelling. If labelling efficiencies for two targets are



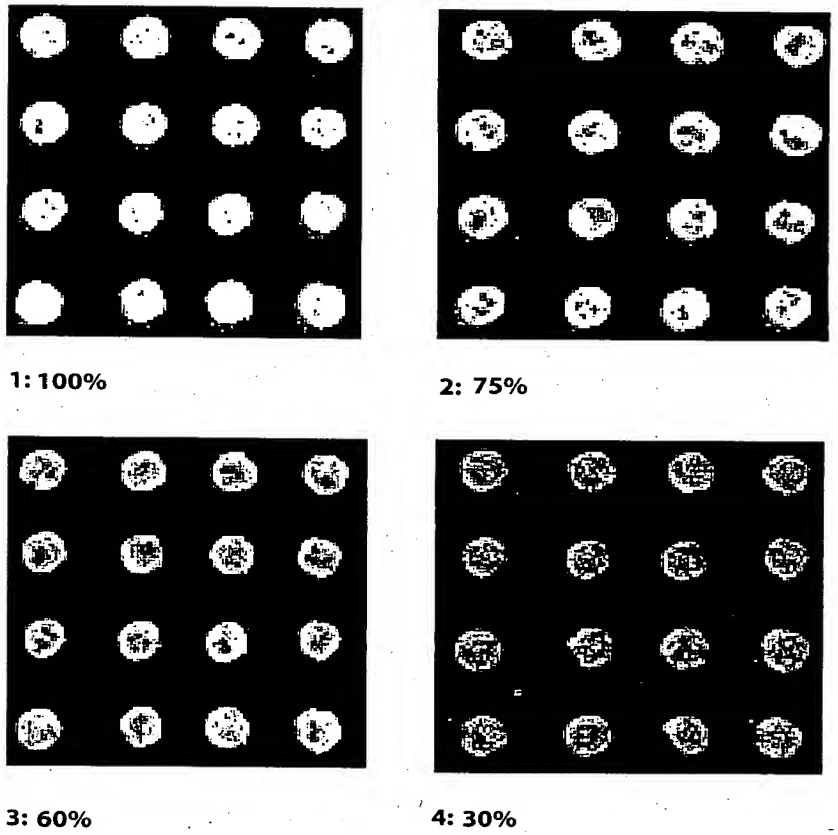



**Fig. 2.2.** Series of images showing amount of Cy3-labelled PCR product attached to a poly-L-lysine-coated substrate at various stages of a microarraying experiment. 1 Freshly arrayed product; 2 after post-array fixation; 3 after denaturation; 4 after overnight hybridisation in 5×SSC/0.1% SDS at 62 °C. All images were scanned at 70% laser power/70% PMT on a Genetic Microsystems 418 array scanner and Imagene 2.0 (BioDiscovery) software was used to quantify amount of element retained

similar, then similar levels of signal are observed after dual hybridisation. RNA quality was assessed by arraying three distinct elements from the *Gapdh* gene: one from the 5′ UTR region of the transcript, another from the open reading frame and a 3′ UTR element. If degradation affects one of the RNA samples, or if reverse transcription was restricted in one of the labelling reactions, this is detected as uneven signals for the 5′ versus the 3′ *Gapdh* elements from the two targets. Note that this control is only suitable when labelling RNA targets by oligo-dT-primed reverse transcription.

The success of hybridisation and post-hybridisation washes can be monitored by spiking the labelled target mixture with a known quantity of fluorescently labelled probe (50 ng) from two bacterial genes: *amp* and *kan*. The bacterial probes can be labelled by incorporation of fluorescent nucleotides during PCR amplification.

3. Key controls comprise those genes (also known as reference genes) that display a known pattern of expression within the system being studied. It is important to include probes for any genes known to display differential expression within the system you are studying, since it is the performance of these genes on the array that act as important references for the behaviour of anonymous cDNA probes. When all the control genes perform appropriately and a panel of known genes displays expected differential expression patterns, data collected from unknown elements can be used with a high level of certainty. A second reason for including probes for known genes relevant to the biological system of interest is that these can act as useful references when examining the expression data using clustering algorithms. Recent advances in data mining have shown it is possible to cluster genes that show similar expression patterns over a series of experiments; in many cases genes which display similar patterns fall into similar functional pathways (Eisen et al. 1998.) Broadly speaking, housekeeping genes fall into this category because they are expected to display constant expression patterns between any test and reference RNA sample. In practice, however, very few genes fail to exhibit some degree of transcriptional modulation in certain contexts.

### 3.2.3 Element Preparation

Element preparation can be separated into the following steps: preparation of template for each element, PCR amplification and element purification. The success of each PCR is determined by gel electrophoresis. Optimal probe concentration is in the order of 200–500 ng/μl. Elements arrayed from more dilute solutions generally give poor signal on hybridisation. The most common source of elements in use today are cDNA clones propagated in bacteria. The inserts from each clone must be amplified by PCR and then



purified from agents that inhibit DNA attachment to the glass substrate. This PCR amplification is typically done using universal vector-based primers to minimise expenditure on amplimers, though gene-specific primers can be used.

The template for these reactions is typically a small aliquot of bacterial culture or purified plasmid DNA. Our experience with the use of cleared bacterial cultures as a template suggests that successful amplification is achieved in approximately 90% of reactions. Amplifying from purified plasmid DNA resulted in very similar levels of success (greater than 90%) but added a considerable amount of effort to progression from elements to arrays. Plasmid preparation did allow for routine sample sequencing of plasmid templates prior to amplification as a guard against errors with plate orientation and plate selection.

PCR product purification can be performed by a variety of methods ranging from column purification (e.g. Qiagen, Telechem, Millipore) to simple isopropanol precipitation. Great care must be taken at each stage to prevent cross-contamination of probes. The preparation of large numbers of bacteria and plasmids in 96-well format and the subsequent amplification of thousands of PCR products by universal vector primers is prone to cross-contamination. Sample sequencing of plasmid templates, diligent use of negative controls and adherence to good PCR technique are imperative to avoid possible disaster.

### 3.2.4 Arrays

Arraying can be broken down into the following aspects: (1) choice of a robotic arraying system; (2) choice of a suitable slide chemistry for DNA attachment; (3) choice of an arraying solution; and (4) post-array fixation.

#### 3.2.4.1 Choosing an Arraying System

The last couple of years have seen a vast expansion in the number of robotic arraying units available on the market. There is a clear trend towards faster, higher throughput microarrayers, with the capacity to print a larger number of slides simultaneously. The delivery system aspect of microarraying has also advanced dramatically and there are now a variety of pin, quill, capillary, piezo-electric and ink-jet spotting systems being used to print DNA elements onto substrates. Time should be taken to assess which system is most appropriate to your needs. Array regularity, spot uniformity, speed and efficiency of spotting are the most critical factors that should be considered when deciding on an arraying system.

### 3.2.4.2 Slide Chemistry

Slide chemistry consists of the way DNA is attached to the non-porous glass substrate and the subsequent inactivation of the substrate post-arraying. When this is done efficiently, DNA elements are successfully bound to the glass surface and then inactivated to prevent labelled cDNA target from binding to the substrate during the hybridisation, resulting in undesirable background. The sensitivity of this technology derives from the ability to detect weak fluorescent signals at a given element on a non-porous substrate in comparison to the very low background surrounding the element.

No matter what arraying system is being used, it is important to determine the best substrate and binding chemistry for a given objective. Some chemistries require the use of modified 5′ amino-linked primers in the PCR amplification used to generate probe. Traditionally, poly-L-lysine- or silane-coated slides have been used. The most popular to date has been poly-L-lysine slides prepared "in house". The advantages of the poly-lysine chemistry are that it requires no DNA modification, it is extremely cheap and, once perfected, it provides a highly consistent performance. The problems associated with home manufacture originate from batch variation in the product, the shelf-life and the poorly understood maturing process that is required between slide coating and microarraying. Commercial poly-L-lysine-coated slides have tended to be substandard for microarray purposes because they were not designed or packaged for this purpose.

More recently, there has been a dramatic increase in the number of commercially available substrates for microarraying. These substrates are generally made using superior glass of uniform thickness and both monolayer (e.g. Corning, Telechem) and branched polymer (e.g. 3D Surmodics) substrates exist. Many of the newer substrates require the use of 5′ amino-modified amplimers for element preparation.

### 3.2.4.3 Array Solution

As previously mentioned, spot consistency is very important for expression profiling. The nature of spotting can be altered by the solution the DNA is resuspended in. Traditional arraying methods involved the spotting of elements in a salt-buffered solution (2–5×SSC), but the use of surfactants and chaotropic agents (e.g. 1 M NaSCN) contributes to a more even deposition of element onto the glass substrate. We employed DMSO (10–25%) in our spotting solution because it promotes denaturation and reduces the evaporation rate of samples to be printed. The selection of an arraying solution is often determined by the chemistry that is used to attach DNA elements to the glass substrate. Also, it is important to remember that spotting solutions will perform differently depending on the element delivery system.



## 3.3 Labelled RNA Targets

One of the most challenging aspects of expression profiling experiments is the efficient and consistent labelling of RNA. If the source of RNA is limited, then linear amplification protocols may be required prior to labelling to ensure adequate amounts of starting material (Kacharmina et al. 1999).

### 3.3.1 Preparing RNA

Differing schools of thought exist on whether total or poly A+ RNA is the best source of RNA template for labelling: each type of RNA may introduce its own particular bias to the experiment. In either case, the RNA must be free of genomic DNA, solvents or other common contaminants retained from simple RNA extractions. The most common cause of problems associated with labelling of RNA targets relates to the quality of the RNA template.

In the case of mammalian tissues that are traditionally known as difficult tissues to extract clean RNA from, muscle for example, caesium banding procedures are recommended. In the case of cell line studies, RNA extraction with guanidinium thiocyanate followed by column purification (e.g. Qiagen *RNAeasy*) is sufficient.

### 3.3.2 Labelling RNA

Traditionally, RNA templates have been labelled by the incorporation of dye-labelled nucleotides into a first strand synthesis reaction using MLV reverse transcriptase (DeRisi et al. 1996). These reactions are extremely inefficient and as a result require high concentrations of dye-labelled nucleotides during labelling, most of which is discarded. A variety of new protocols have been adapted to circumvent wastage of dye-labelled nucleotides, including recycling of unincorporated label by HPLC (see http://cmgm.stanford.edu/pbrown/mguide/hplc.html).

Recently, the non-covalent attachment of fluorophores to nucleic acid involving a chemical reaction instead of the traditional enzymatic process has become commercially available (http://www.probes.com/media/pis/mp21650.pdf). Also, efforts have been put into the generation of labelling systems allowing cDNA synthesis using modified nucleotides, which incorporate far more efficiently than the traditional Cy-dye-labelled dNTPs. Fluorescent dyes are then ester linked to free amino groups present on the modified nucleotides incorporated into the first strand cDNA(http://

cmgm.stanford.edu/pbrown/protocols/aadutpcouplingprocedure.html). Such labelling methods are substantially cheaper than the traditional labelling protocols given the fact that mono-reactive Cy dyes are one-tenth the cost of Cy-dye-labelled nucleotides.

When RNA samples are limiting, there are several alternatives: protocols that combine RNA target amplification, efficient labelling and signal amplification have also been commercialised (e.g. MicroMax NEN Dupont). Many of these new methodologies are in their early stages of development so it is still unclear how well they compare to the traditional labelling method. Another recent advance has seen the use of dye-labelled-oligo dendrimers to amplify hybridisation signals (Genosphere, http://www.genisphere.com).

A final comment concerning labelling procedures is that good data are dependent on attaining the best signal-to-background ratio possible. Efficient removal of unincorporated products, avoiding evaporation of hybridisation solution while on the microarray and cleanliness of reagents therefore must be a priority.

### 3.3.3 Hybridisation

Hybridisation conditions for microarrays classically mimic those of other molecular methods. Stringency is controlled by using reagents or parameters common to liquid hybridisation solutions (e.g. SSC, formamide and temperature control). Non-specific background is also prevented by using familiar agents such as SDS and Denhardt's solution. All hybridisation solutions contain large amounts of blocking agents designed to minimise hybridisation due to the presence of repeat sequences or poly A tracts on labelled cDNAs.

The very small volumes used for hybridisation make evaporation a critical factor in experimental failure. Evaporation leads to areas of local background where the hybridisation solution dries onto the surface of the array and cannot be removed by washing. Hybridisation chambers are a common tool used to reduce the chance of evaporation and thus avoid ruining an experiment. The array is hybridised with the labelled target under a coverslip and sealed into a shallow, humidified chamber.

Bubbles need to be avoided when applying coverslips as they lead to local variations in background. The two most common ways of applying the hybridisation solution to the array are to (1) use capillary action to draw the hybridisation solution under the coverslip or (2) deposit the hybridisation solution onto the array and overlay the coverslip. Poor coverslip placement can lead to wicking of the hybridisation solution down the side of the slide and into the chamber.



### 3.3.4 Scanning and Image Analysis

Microarray readers can be broken into two major classes: confocal laser scanners and CCD (charged couple device) camera-based devices. Each class has its own advantages and disadvantages. CCD systems are often more affordable since the mode of illumination is cheaper and the fluorescent signal can be collected over long periods of time. Such systems also are amenable to use of a wide variety of fluorophores due to the wide excitation spectrum of their illumination sources. This utility comes at a price, however, since these systems traditionally have had reduced signal sensitivity, problems with non-specific signal discrimination, and poor resolution of the collected image.

The confocal laser scanning approach gives excellent sensitivity and high resolution (5–10 μ as standard). In some cases multiple lasers are now used for dual colour detection system and image analysis.

### 3.3.5 Validation of Leads

In order to guarantee the production of reliable data sets it is best to replicate microarray experiments. It is currently recommended that four sets of data should be generated from each element. When test and reference RNAs are not in limiting amounts, four independent experiments are performed. Alternatively, duplicate experiments are sufficient if each element has been spotted in two positions on the array.

Several studies have used independent methods to accurately validate expression profiles observed with microarrays, using quantitative assays, e.g. quantitative RT-PCR and Northern blotting. Because the majority of our microarray experiments have focused on expression in whole developing mouse organ systems, we chose whole-mount in situ hybridisation for independent validation of our data. Whole-mount in situ hybridisations allowed us to confirm differential expression between organs and also defined cell- or tissue-specific expression in the organ of interest.

## 4 Conclusions

In this chapter we have tried to highlight some of the issues that are critical for the design and implementation of expression profiling experiments using cDNA microarrays. There are several topics that have not been covered, such as approaches to data normalisation and data mining, but these are dealt with elsewhere in this volume.

It is important to emphasise a flexible approach when using microarrays: utilise those resources and conditions that meet your needs. This technology is developing at such a fast rate that many of the approaches described here will probably be superseded in the coming months and years. It is important that researchers take time to develop protocols and try new approaches, rather than slavishly following published protocols. The protocols included here are meant to serve as a guide, and no more.

In our opinion, some variant of the approach described here is likely to become as important to basic biological research, as well as clinical research, as PCR methodology has been in the last decade. Whether developmental biology or toxicology, most areas of research will benefit enormously from the insights offered by expression profiling of complete genome sets in the near future.

## 5 Protocols

The protocols described here have been used by us routinely for expression analysis in the developing mouse embryo (Grimmond et al. 2000). They were employed in conjunction with hardware from GMS/Affymetrix and software from BioDiscovery. They are derivations of protocols of Schena and Davis, the lab of P. Brown (Stanford) and Hegde et al. (2000). The original protocols can be found at http://arrayit.com/dna-microarray-protocols and http://cmgm.stanford.edu/pbrown/protocols/index.html.

### 5.1 Poly-L-Lysine-Coated Substrate Preparation

We follow the protocol from the Brown lab web site(http://cmgm.stanford.edu/pbrown/protocols/1_slides.html).

### 5.2 Element Preparation

#### 5.2.1 Plasmid Preparation

Elements are normally prepared using an alkaline lysis method that is performed in 96-well format using multi-channel pipettes. Generally, the DNA is of good enough quality for sequencing.

1. Aliquot inoculate 1.1 ml of LB broth + ampicillin 50 µg/ml into 2.2-ml-deep 96-well plates. (Plates with square wells and round bottoms are best.)



Cover the plate with sealable lid and shake at 250 rpm (minimum) overnight at 37 °C.
2. Spin down bacterial pellet by spinning plate at 2,000 *g* for 15 min, 4 °C.
3. Tip off supernatant and pat dry on towel.
4. Resuspend bacterial pellet in 100 µl of buffer P1 (10 mM Tris pH 8.0, 1 mM ETDA pH 8.0, 100 µg/ml RNase A and RNase T1. Leave for 5 min at ambient temperature.
5. Add 100 µl of buffer P2 (0.2% SDS, 0.2 M NaOH). Leave on ice for 5 min.
6. Add 100 µl of 3 M Na acetate pH 4.8 and precipitate at –20 °C for 30 min.
7. Spin plate at 3,000 *g* for 60 min at 4 °C and transfer 125 µl to a new 96-well plate. Add 125 µl of isopropanol and precipitate at –20 °C for at least 1 h.
8. Precipitate plasmid DNAs by spinning plate at 3,000 *g* for 60 min at 4 °C. Tip out supernatant. Wash with 500 µl of 70% ethanol. Spin at 3,000 *g* for 10 min at 4 °C. Leave plate for 60 s to dry on bench.
9. Resuspend in 50 µl of TE and run 5 µl on a gel to check the preparation. Dilute 10 µl down to 2 ng/µl for generating PCR products.

#### 5.2.2 PCR Amplification

1. PCR reactions are set up as follows:

| Reagent | Per reaction | Per 4×96 reactions |
|---|---|---|
| 10×PCR buffer | 10 | 4.8 ml |
| 25 mM $MgCl_2$ | 7 | 4.0 ml |
| 100 mM $NH_2$-M13F* | 0.3 | 140 µl |
| 100 mM $NH_2$-M13R* | 0.3 | 140 µl |
| 5 U/µl Taq (AB Biotech) | 0.8 | 320 µl |
| $H_2O$ to | | 48 ml |

The primers (*) have a 5′ $NH_2$ linker group with a $C_6$ spacer. The primer sequences (17) are:

$NH_2$-M13F $NH_2$-C6-GTT TTC CCA GTC ACG AC
$NH_2$-M13R $NH_2$-C6-ACA GGA AAC AGC TAT GAC

2. Dispense bulk mix into a standard profile polypropylene 96-well plate (Thermofast, AB Biotech). Low profile plates are not suitable since they cannot hold the volumes required for isopropanol precipitation of the PCR products.
3. Add 5 µl of plasmid DNA to each reaction.
4. PCR reactions are carried out in an MJ research tetrad PCR machine under the following conditions: (temperature of block is calculated, plate option is selected, volume of reaction selected is 100 µl in the program setup) 1×94 °C for 180 s, 35×(48 °C for 20 s, 72 °C for 180 s, 94 °C for 40 s).
5. Run 5 µl of the PCR reaction on a 96-lane gel to determine success of the PCR amplification. Record failures and then re-amplify corresponding samples.

### 5.2.3
### Product Purification

1. Immediately after electrophoresis, add 10 µl of 3 M Na acetate (pH 5.2) and 100 µl isopropanol. Seal the plate and precipitate at –20 °C overnight. Elements are recovered by spinning at a minimum of 3,000 rpm for 60 min at 4 °C. Invert plates onto an absorbent towel and gently tap dry. Pellets are often just visible.
2. Wash pellets with 500 µl of 70% ethanol. Spin at 3,000 rpm for 10 min at 4 °C and blot plates dry on fresh towel.
3. Leave to air dry for 60 min to overnight. Seal plates and store at –20 °C until ready to array.
4. Prior to arraying, resuspend pellets in the desired array solution.

## 5.3
## Post-Array Fixation

Once again, this protocol was adapted from the Brown web site (http://cmgm.stanford.edu/pbrown/protocols/3_post_process.html).

## 5.4
## RNA Target Labelling

We have tried all sorts of variations on the theme of reverse transcriptase (RT) incorporation of Cy-dyes into first strand cDNA. We have performed this on mRNA prepared by a variety of kits (Invitrogen, Pharmacia and Qiagen), and when it did work, the Qiagen kit was probably the best. RNA preparation is a *critical* step for ensuring good results. In our hands, results are entirely dependent on RNA quality. Anything less than perfect and we generally saw no signal.

More recently, we have begun to use total RNA. We have used the Qiagen RNeasy kit and seen an improvement in the robustness of our labelling experiments. Note: this works well on embryonic tissues that lyse very easily and contain little connective tissue at the stages at which we work. The use of adult tissues may require far more diligent preparation (e.g. homogenisation in GITC and CsCl banding).

RNA is prepared using RNeasy protocol. A Qiashredder is incorporated to minimise gDNA contamination. We also check the purified RNA with primers from the 3′ UTR of *Gapdh* in a control RT-PCR experiment to confirm the absence of genomic DNA. If genomic DNA is present, the RNA should be treated with RNase-free DNase for 30 min at 37 °C followed by inactivation of the DNase at 70 °C for 10 min. RNA is made just prior to labelling when possible, or stored at –70 °C if necessary.



The following protocol has been used routinely with 20 µg of total RNA. Up to 100 µg of total can be used; 1–5 µg of mRNA can also be used.

1. Mix the following:

| | |
|---|---|
| RNA (10 µg/µl) | 2 µl |
| Oligo-dT ($T_{18-25}$) (2 µg/µl)* | 2 µl |
| $H_2O$ | 20 µl |

2. Heat to 70 °C for 10 min. Cool to 37 °C by putting into preset hot block for 10 min.
3. Prepare the following:

| | |
|---|---|
| 5×RT buffer (BRL) | 10 µl |
| DTT | 5 µl |
| 25.0 mM dATP, dTTP, dGTP | 1 µl |
| 2.5 mM dCTP | 2 µl |
| 1.0 mM Cy3 or Cy5 dCTP | 4 µl |
| RNA-Guard | 1.5 µl |

4. Add 23 µl of the above mix to the RNA + oligo dT mix. Spin down with 10-s pulse in the centrifuge and replace at 37 °C for 1 min. Add 2 µl of RTase (Superscript II, BRL Life Technologies). Incubate for 90 min at 37 °C.
5. Add 1 µl of 0.5 M EDTA and 2 µl of 2 M NaOH. Heat to 65 °C for 10 min. Chill on ice for 1 min and add 4 µl 1 M HCl and 4 µl 1 M Tris pH 8.0.
6. Remove unincorporated dye-labelled nucleotides by ethanol precipitation. Add 10 µl of 1 µg/µl Cot1 DNA (Life Technologies), 7 µl of 3 M Na acetate (pH 4.8) to the reaction. Mix well. Add 150 µl of ethanol and chill at –20 °C for 30 min. Spin down the pellet (14,000 rpm, 4 °C, 20 min) and wash pellet well with 500 µl of 70% ethanol. Allow pellet to air dry but do not over dry.
7. Resuspend both pellets in the same 50 µl of microarray hybridisation solution (4×SSC/0.2% SDS/50% formamide).
8. Add 1 µl of hybridisation standards and heat to 80 °C for 10 min to denature probe. Cool on ice and add 1 µl 10 mg/µl poly dA (Pharmacia). Incubate the labelled target at 42 °C for at least 1 h prior to placing on array. This competes out repeat sequences and poly T tracts.
9. Cool hybridisation mix on ice for 1 min (this prevents a halo from hot hybridisation solution drying on the slide). Spin for 1 min at maximum speed to collect condensation and to precipitate any possible particulate material.
10. Place hybridisation solution onto array, avoiding bubbles! This is best done by carefully laying the coverslip onto slide with fine forceps. Alternatively, slowly pipette the hybridisation solution next to a coverslip placed over a dry slide. Capillary action draws the hybridisation solution under the coverslip. Applying hybridisation solution is worth practising, especially with the large 60-mm coverslips.
11. Place slide into hybridisation chamber. Add 10 µl of hybridisation mix into the wells at each end of the chamber. Tighten lid and submerge in 45 °C water bath for 14–24 h.

12. After hybridisation, dry chamber, place slide with coverslip into slide holder and plunge into 1 l of 0.2×SCC, 0.05% SDS at ambient temperature until coverslip falls off. Rock for 3 min. Rapidly transfer slide holder into a second beaker of 1 l 0.2×SSC and rock for a further 3 min.
13. Carry beaker plus slides to centrifuge set at ambient temperature. Slides can be conveniently spun in a 50-ml plastic V-bottomed tube if racks are not available. Remove slide with forceps and spin at 500 rpm for 3 min. If possible, put the "handled" end of slide into the carrier so that is spins on the outside during centrifugation. This prevents contaminants getting onto the array.

## 5.5 Controls

### 5.5.1 Sensitivity Controls

We use several sets of controls: the first set is a series of pre-labelled PCR products that are added at known concentrations to each hybridisation. These labelled products allow the efficiency of hybridisation to be gauged.

| | |
|---|---|
| *Amp* PCR product | 50 ng/hyb |
| *Kan* PCR product | 5 ng/hyb |

1. Products are labelled by incorporating Cy dyes during amplification.
2. For a 25-µl reaction:

| | |
|---|---|
| 10×PCR buffer (plus $MgCl_2$) | 2.5 µl |
| Low CTP mix (2 mM A, G, TTP, 0.4 mM dCTP) | 2.5 µl |
| Cy3 or Cy5 CTP (1 mM) | 2.0 µl |
| 10 µM oligos | 2.0 µl |
| Taq (5 U/µl) | 0.5 µl |
| $H_2O$ to | 25 µl |

3. Prepare a Cy3- and Cy5-labelled product for each control gene. Use standard amplification procedures. Remove unincorporated Cy-CTP by Qiaquick PCR product purification column.
4. Dilute Cy3 and Cy5 products for each control gene to the same concentration. Make up stock tube that contains labelled gene products at the following concentrations: Cy3 and Cy5 *Kan* products, 50 ng/µl; Cy3 and Cy5 *Amp* products, 5 ng/µl; and Cy3 and Cy5 Plant1 gene products, 500 pg/µl.

### 5.5.2 Labelling Controls

To our target RNA we add in vitro RNA generated from a series of plant genes known to contain large poly A tails. The non-mammalian RNA is added prior



to each reaction to determine the efficiency of the RT step. RNA is made using the Ambion RNA megascript kit. RNA samples are diluted out to the following concentrations and added to hybridisations:
Plant gene 1 in vitro RNA:50 ng/hyb
Plant gene 2 in vitro RNA: 5 ng/hyb

### 5.5.3 RNA Integrity Control

We also print three *Gapdh* PCR products onto each array. These products are derived from the 5′ UTR, ORF and 3′ UTR regions of the *Gapdh* transcript. If RNA is degraded in one of the samples or if the RT labelling has generated truncated products, all three elements will not give a strong co-expressed signal.

## References

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Staudt LM (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403(6769):503–511

Bonaldo MF, Lennon G, Soares MB (1996) Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res 6(9):791–806

DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 14(4):457–460

DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278:680–686

Dixon AK, Richardson PJ, Lee K, Carter NP, Freeman TC (1998) Expression profiling of single cells using 3 prime end amplification (TPEA) PCR. Nucleic Acids Res 26(19):4426–4431

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95(25):14863–14868

Grimmond S, Van Hateren N, Siggers P, Arkell R, Larder R, Soares MB, Bonaldo M, Smith L, Tymowska-Lalanne Z, Wells C, Greenfield A (2000) Sexually dimorphic expression of protease nexin-1 and Vanin-1 in the developing mouse gonad prior to overt differentiation suggests a role in mammalian sexual development Hum Mol Genet 9(10):1553–1560

Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Earle-Hughes J, Snesrud E, Lee N, Quackenbush J (2000) A concise guide to cDNA microarray Anal Biotech 29(3):548–562

Kacharmina JE, Crino PB, Eberwine J (1999) Preparation of cDNA from single cells and subcellular regions. Methods Enzymol 303:3–18

Luo L, Salunga RC, Guo H, Bittner A, Joy KC, Galindo JE, Xiao H, Rogers KE, Wan JS, Jackson MR, Erlander MG (1999) Gene expression profiles of laser-captured adjacent neuronal subtypes. Nat Med 5(1):117–122

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470

Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci USA 93:10614–10619

Soares MB, Bonaldo MF, Jelene P, Su L, Lawton L, Efstratiadis A (1994) Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci USA 91(20):9228–9232